

Limit Order Strategic Placement with Adverse Selection Risk and the Role of Latency.

Charles-Albert Lehalle* and Othmane Mounjid†

Printed the December 5, 2016

Abstract

This paper is split in three parts: first we use labelled trade data to exhibit how market participants accept or not transactions via limit orders as a function of liquidity imbalance; then we develop a theoretical stochastic control framework to provide details on how one can exploit his knowledge on liquidity imbalance to control a limit order. We emphasize the exposure to adverse selection, of paramount importance for limit orders. For a participant buying using a limit order: if the price has chances to go down the probability to be filled is high but it is better to wait a little more before the trade to obtain a better price. In a third part we show how the added value of exploiting a knowledge on liquidity imbalance is eroded by latency: being able to predict future liquidity consuming flows is of less use if you do not have enough time to cancel and reinsert your limit orders. There is thus a rationale for market makers to be as fast as possible as a protection to adverse selection. Thanks to our optimal framework we can measure the added value of latency to limit orders.

To authors' knowledge this paper is the first to make the connection between empirical evidences, a stochastic framework for limit orders including adverse selection, and the cost of latency. Our work is a first step to shed light on the roles of latency and adverse selection for limit order placement, within an accurate stochastic control framework.

1 Introduction

With the electronification, fragmentation, and increase of trading frequency, orderbook dynamics is under scrutiny. Order flow dynamics is of paramount importance since it plays a major role in price formation. At the smallest time scale order placements have been studied as a system using economics, econophysics and applied mathematics (see [Lo et al., 2002], [Bouchaud et al., 2004], [Huang et al., 2015b] and references herein). On the one hand, at a larger time scale, investors' decisions are split into large collections of limit (i.e. liquidity providing) orders and market (i.e. liquidity consuming) orders and contribute to price formation (see [Kyle, 1985], [Tóth et al., 2012], [Bacry et al., 2014] and their references for three different viewpoints on price formation and market impact). The two time scales are linked since dynamics around small orders are shaping the market impact of the large metaorders they are part of (see [Zarinelli et al., 2015] for a discussion about this relationship).

*Capital Fund Management, Paris and Imperial College, London

†Université Pierre et Marie Curie, Paris

On the other hand, (high frequency) market makers mostly use limit orders, providing liquidity to child orders of investors. Price dynamics around their orders have been studied too (see for instance [Biais et al., 2016] for small time scales and [van Kervel and Menkveld, 2014] for large time scales).

In practice market makers and investors use optimal trading strategies to bind the two time scales. Models for these strategies are now well known (see [Guéant et al., 2013] and [Guéant et al., 2012] for a common framework involving limit orders for market makers and investors). For obvious reasons, the focus of papers about optimal trading strategies has been risk control. In practice market participants combine short term anticipations of price dynamics inside these risk control framework ([Almgren and Lorenz, 2006] for the inclusion of a Bayesian estimator of the price trend in a mean-variance optimal trading strategy or [?] for an inclusion of estimates of future liquidity consumption – μ in their paper should be compared to our consuming intensities λ^{\pm} – in macroscopic optimal execution.).

This paper focuses on the optimal control of one limit order (potentially included in an optimal strategy at a largest time scale, see [Lehalle et al., 2013, Chapter 3] for a practitioner viewpoint on splitting the two time scales of metaorders executions) taking profit of a short term anticipation of price moves.

After some considerations about short time predictions in orderbooks and empirical evidences showing market participants take imbalance into account, we first show that in the context of a very simple choice (canceling or (re) inserting a limit order) optimal control can add value to any short term predictor. This result can thus be used by investors or market makers to include some predictive power in their optimal trading strategies.

Then we show how latency influences the efficient use of such predictions. As expected the more latency to take decision, the less one can take profit of optimal control. It allows us to link our work to regulatory questions. First of all: what is the “value” of latency? Regulators could hence rely on our results to take decisions about “slowing down” or not the market (see [Fricke and Gerig, 2014] and [Budish et al., 2015] for discussions about this topic). It sheds also light on maker-taker fees since the real value of limit orders (including adverse selection costs), are of importance in this debate (see [Harris, 2013] for a discussion).

This paper can be seen as a mix of two early works presented at the “Market Microstructure: Confronting Many Viewpoints” conference (Paris, 2014): a data-driven one focused on the predictive power of orderbooks [Stoïkov, 2014], and an optimal control driven one [Moallemi, 2014]. Our added values are first a proper combination of the two aspects (inclusion of an imbalance signal in an optimal control framework for limit orders), and then the construction of our cost function. Unlike in the second work, we do not value a transaction with respect to the mid-price at $t = 0$, but with respect to the microprice (i.e. the expected future price given the liquidity imbalance) at $t = +\infty$. We will argue the difference is of paramount importance since it introduces an effect close to adverse selection aversion, that is crucial in practice.

As an introduction of our framework, we will use a database of labelled transactions on NASDAQ OMX (the main Nordic European regulated markets) to show how orderbook imbalance is used by market participants in a way that can be seen as compatible with our theoretical results.

Hence the structure of this paper is as follows: Section 2 presents orderbook imbalance as a microprice and details our optimal control framework, and Section 3 illustrates the use of imbalance by market participants thanks to the NASDAQ OMX database. Once these elements are in place, Section 4 presents our model and 5 shows how to numerically solve the control problem and provides main results, especially the influence of latency on the efficiency of the strategy.

2 The Model

2.1 The Orderbook Imbalance as a Microprice

Martingality of the price changes no more stands once you add information to your knowledge of the state of the “world” (i.e. the *filtration* supporting the randomness of your perception of price dynamics). At small time scales, well known stylized facts break the martingality of price changes (see [Cont, 2001] or references inside [Lehalle, 2014]) like the positive autocorrelation of the signs of trades (i.e. there is more chance the next transaction is initiated by a buyer or a seller if the previous ones have been too) and the negative autocorrelation of the returns (once the price moved up –respectively down– the probability it goes down –resp. up– is larger than if the price went down –resp. up–).

When you add the state of the orderbook to your filtration, next prices moves are even “less martingale” (i.e. easier to predict). Academic papers (like [Bouchaud et al., 2004] or [Huang et al., 2015b]) and brokers’ research papers (see [Besson et al., 2016]) document how the sizes at first limits of the public orderbook¹ influences the next price change.

It is worthwhile to underline the identified effects are usually not strong enough to be the source of a statistical arbitrage: the expected value of buying and selling back using accurate predictions based on sizes at first limits does not beat transaction costs (bid-ask spread and fees). See [Jaisson, 2014] for a discussion. Nevertheless

- for an investor who *already took the decision to buy or to sell*, this information can spare some basis points. For very large orders, it makes a lot of money and in any case it reduces implicit transaction costs.
- market makers naturally use this kind of information to add value to their trading processes (see [Fodra and Pham, 2013] for a model supporting a theoretical optimal market making framework including first limit prices dynamics).

The easiest way to summarize the state of the orderbook without destroying its informational content is to compute its *imbalance*: the quantity at the best bid minus the one at the best ask divided by the sum of these two quantities:

$$\text{Imb}_t := \frac{Q_t^{\text{Bid}} - Q_t^{\text{Ask}}}{Q_t^{\text{Bid}} + Q_t^{\text{Ask}}}. \quad (1)$$

The future price move is positively correlated with the imbalance. In other terms

$$\mathbb{E}((P_{t+\delta t} - P_t) \times \text{sign}(\text{Imb}_t) | \text{Imb}_t) > 0, \quad (2)$$

where P_t is the midprice (i.e. $P_t = (P_t^{\text{Bid}} + P_t^{\text{Ask}})/2$, where P_t^{Bid} and P_t^{Ask} are respectively the best bid and ask prices) at t for any δt . Obviously when δt is very large, this expected price move is very difficult to distinguish from large scale sources of uncertainty. See for instance [Lipton et al., 2013] for details on the “predictive power” of such an indicator (our Figure 1 illustrates this predictive power on real data).

The nature of the predictive power of the imbalance. It is outside of the scope of this paper to discuss and document the predictive power of the imbalance; we just give here some clues and intuitions to the reader:

¹Limit orderbooks are used in electronic market to store unmatched liquidity, the *bid size* is the one of passive buyers and the *ask size* the one of passive sellers; see [Lehalle et al., 2013] for detailed.

- first of all, since the quantity at the bid is the unmatched buying quantity and the one at the ask the unmatched selling one, it is natural to deduce at one point owners of the associated orders will loose patience and consume liquidity, pushing the price in their direction.
- within a model in which market orders occurrence follow independent point processes of the same intensity, the smallest queue (bid or ask) will be consumed first, and the price will be pushed in its direction. See [Huang et al., 2015b] for a more sophisticated point process-driven model and associated empirical evidences.
- Another viewpoint on imbalance would be the bid vs. ask imbalance contains information about the direction of the net value of investors' metaorders. Or directly (if one is convinced investors post limit orders), or indirectly (if one believe investors only consume liquidity and in such a case bid and ask sizes are an indicator of market makers net inventory).

The focus of formula (1) on two first limits weaken the predictive power of the bid vs. ask imbalance. For large tick assets² it may be enough to just use the first limits, but for small tick ones it certainly increases the predictive power of our imbalance indicator to take more than one tick into account. Since a discussion on the predictive power of imbalance is outside the scope of the paper, we will stop here the discussion.

This paper is providing a stochastic control framework to post limit orders using the information contained in the orderbook imbalance. In such a context, we will call the *microprice seen from t* and note $P_{+\infty}(t)$:

$$P_{+\infty}(t) = \lim_{\delta t \rightarrow +\infty} \mathbb{E}(P_{t+\delta t} | P_t, \text{Imb}_t). \quad (3)$$

2.2 Construction of an Associated Optimal Control Framework

To setup a discrete time framework for optimal control of one limit order with imbalance, we focus on the simple case (but complex enough in terms of modelling) of one atomic quantity q to be executed in T_f units of time (can be orderbook events, trades, or seconds). It will be a buy order, but it is straightforward for the reader to transpose our result for a sell order.

From zero to T_f the trader (or software) in charge of this limit order can cancel it (i.e. remove it from the orderbook), insert it at the top of the bid queue (if it is not already in the book), or do nothing. The natural dynamics of the system are driven by point processes consuming and filling the two sides of the orderbook, conditionally to its imbalance: $N^{Opp,+}$ and $N^{Opp,-}$ for addition and consumption on the *opposite side* of the book (i.e. the best ask in the case of a buy order), and $N^{Same,+}$ and $N^{Same,-}$ on the *same side* of the book (i.e. best bid in our case). It will allow us to write a very straightforward way the state of the two first queues, keeping in mind we have to keep track of two quantities on the bid side: the quantity *before the order* Q^{Before} (i.e. to be executed before our limit order is first in the queue) and the one *after the order* Q^{After} , while one quantity is enough on the opposite side: Q^{Opp} . We will use the notation $Q^{Same} := Q^{Before} + Q^{After}$ for the quantity at the best bid, and by convention we will write $Q^{After} = 0$ when our order is not in the queue.

When one of the two queues fully depletes, we will assume the price will move in its direction, and we will need a *discovered quantity* Q^{disc} to replace the deleted first queue, and an *inserted quantity* Q^{ins} to be put in front of the not depleted queue. These quantities will be random variables and their law will be conditioned by the imbalance before the depletion.

²For a focus on tick size, see [Dayri and Rosenbaum, 2015].

For simplicity, we will assume decrease of queues will be caused by transactions (and not by cancellation). This would not change dramatically our framework to consider cancellations, but the writing of the dynamics will be far more complex.

Terminal state. In our framework, if the trader did not obtain an execution thanks to its optimal positing policy at T_f , we force him to cancel his order (if any) and to send a market order to obtain a trade.

Choice of a benchmark. We decide to value the position (i.e. the bought shares) at the best bid price for a limit order or at the best ask for a market order (eventually at $t = T_f$). We compare the value of the obtained shares at t to *its expected value at infinity*, i.e. $\mathbb{E}(P_{+\infty}(t)|\text{Imb}_t)$. It is important since, thanks to this framework

- it is not attractive to buy at the best bid if we expect the price to continue to go down, because it is possible to expect better future price moves thanks to the observed imbalance.
- It thus induces an *adverse selection cost* in the framework, despite we will use the expectation of a linear value function.

This is not a detail since the *trader will have no incentive to put an limit order at the top of a very small queue if the opposite side of the book is large*. We will see in Section 3 using empirical evidences that this is a realistic behaviour. Such a behaviour cannot be captured by other linear frameworks like [Moallemi, 2014].

3 Main Hypothesis and Empirical Evidences

The predictive power of the orderbook imbalance is well known (see [Lipton et al., 2013]). The rationale of this stylized fact (i.e. *the midprice will go in the direction of the smaller size of the book*) is outside of the scope of this paper. We will essentially use the following stylized mechanism: for most stocks, the sizes of the two first limits³ will decrease so that, on average, the smallest one will deplete first leading to a midprice move in its direction. Moreover: with no exogenous information, the discovered limit on the depleted side is now on average larger than the just inserted quantity on the opposite side.

Figure 1 shows the imbalance (1) on the x -axis and the midprice move after 50 trades on the y -axis. The data used here are a direct feed on NASDAQ-OMX, that is the primary market⁴ on the considered stock. Capital Fund Management feed recordings for AstraZeneca reports NASDAQ-OMX accounts for 72% of market share (in traded value) for the continuous auction on this stock over the considered period. Surprisingly, there is currently no academic paper comparing the predictive power of imbalances of different trading venues on the same stock. It is outside of the scope of this paper to elaborate on this. We will hence consider the liquidity on our primary market is representative of the state of the liquidity on other “large” venues (namely Chi-X, BATS and Turquoise on the considered stocks). If it is not the case it will nevertheless not be difficult to adapt our result relying on statistics on each venue, or on the aggregation of all venues. We did not aggregated venues ourselves for obvious synchronization reasons: we do not know the capability of each market participant to synchronize information coming from all venues and do not want to add noise by making more assumptions. Our idea here is to use the state of liquidity at the first limits on the primary market as a proxy of information about liquidity really used by participants .

³We focus on the first limits, i.e. dedicated for *large tick* stocks (see [Huang et al., 2015a] for details) but the reasoning would be the same with an arbitrary aggregation of several ticks for smaller ticks.

⁴i.e. the *regulated exchange* in the MiFID sense, see [Lehalle et al., 2013] for details.

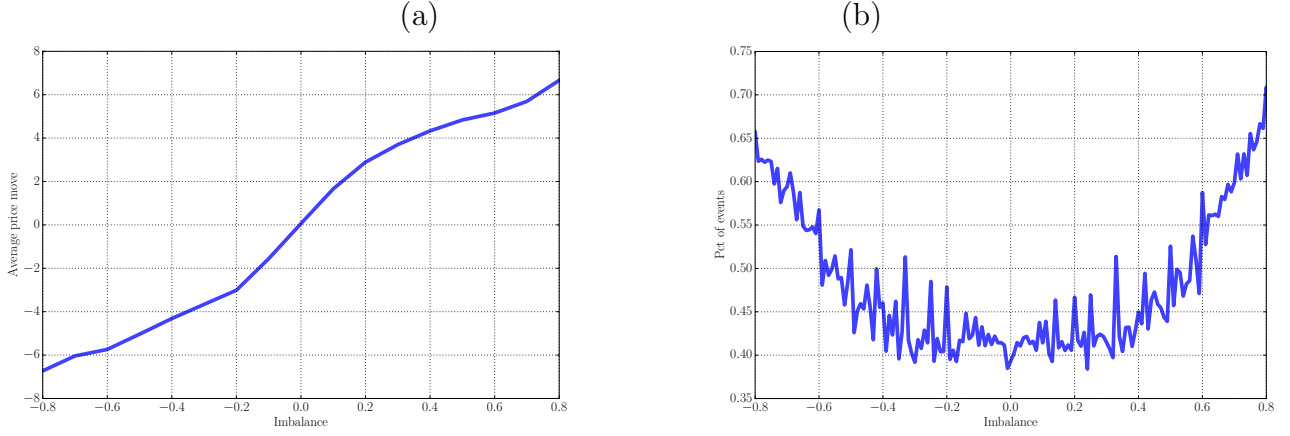


Figure 1: (a) The predictive power of imbalance on stock Astra Zeneca: imbalance (just before a trade) on the x -axis and the expected price move (during the next 50 trades) on the y -axis. (b) distribution of the imbalance just before a trade. From the 2013-01-02 to the 2013-09-30 (accounts for 376,672 trades).

Venue	AstraZeneca	Vodafone
BATS Europe	7.16%	7.63%
Chi-X	19.27%	20.02%
Primary market	72.24%	61.09%
Turquoise	1.33%	11.26%

Table 1: Fragmentation of AstraZeneca (compared to Vodafone) from the 2013-01-02 to the 2013-09-30

We focus on NASDAQ-OMX because this European market has an interesting property: market members' identity is known. It implies transactions are labeled by the buyer's and seller's names. Almost all trading on NASDAQ Nordic stocks was labelled this way until end of 2014 (more details are available in [van Kervel and Menkveld, 2014], because this whole paper is based on this labelling). Note members' identity is not investors' names; it is the identity of brokers or market participants large enough to apply for a membership. High Frequency Participants (HFP) are of this kind. Of course some participants (like large asset management institutions) use multiple brokers, or a combination of brokers and their own membership. Nevertheless, one can expect to observe different behaviours when members are different enough. We will here focus on three classes of participants: High Frequency Participants (HFP), global investment banks, and regional investment banks.

The main hypothesis of this paper is some market participants take the state of the order-book into account to accept or not a transaction. We have in mind one participant can invest to have a good overview of the microscopic state of the liquidity before inserting or cancelling limit orders, or before sending a marketable order. On the principle, academic papers have shown the predictive power of bid-ask imbalance. In this paper, we try to understand if it is used by some market participants, and how to use it.

We will show how optimal control can provide an efficient way to take such information into account. Moreover, latency is of paramount importance when one want to take into account the microscopic state of liquidity: one can be perfectly aware of liquidity imbalance and know how to use it optimally, but can be prevented to do so because of his latency to matching engines of exchanges. The lower frequency of a participant, the less number of times he will be able to take value-adding decisions between $t = 0$ and $t = T_F$.

Main hypothesis of this paper: exploitation of imbalance for limit orders. We expect some market participants to invest in access to data and technology to be able to take profit of the informational content of orderbook imbalance. A very simple way to test this hypothesis is to look at orderbook imbalance just before a transaction with a limit order of a given class of participant. We will focus on three classes of *agents* (i.e. market participants): High Frequency Participants (HFP), Global Investment Banks or Brokers, and Regional Investment Banks or Brokers. Table 2 provides descriptive statistics on these classes of participants in the considered database.

Order type	Participant type	Order side	Avg. Imbalance	Nbe of events
Limit	Global Banks	Sell	0.35	62,111
		Buy	-0.38	63,566
	HFP	Sell	0.32	52,315
		Buy	-0.33	46,875
	Instit. Brokers	Sell	0.57	6,226
		Buy	-0.52	4,646

Table 2: Descriptive statistics for our three classes of agent. AstraZeneca (2013-01 to 2013-09).

We focus on limit orders since information processing, strategy and latency play a more important role for such orders than for market orders (market orders can be sent blindly, just to finish a small metaorder or to cope with metaorders late on schedule, see [Lehalle et al., 2013] for elaborations on brokers’ trading strategies).

For the following charts, we use labelled transactions from NASDAQ-OMX⁵ and thanks to timestamps (and matching of prices and quantities) we synchronize them with orderbook data (recorded from direct feeds by Capital Fund Management). It enables us to snapshot the sizes at first limits on NASDAQ-OMX just before the transaction.

Say for a given participant (i.e. *agent*) a the quantity at the best bid (respectively best ask) is $Q_\tau^{Bid}(a)$ (resp. $Q_\tau^{Ask}(a)$) just before a transaction at time τ involving a limit order owned by a . We note $Q_\tau^{same}(a) := Q_\tau^{Bid}(a)$ (respectively $Q_\tau^{opposite}(a) := Q_\tau^{Ask}(a)$) if it was a buy limit order, or $Q_\tau^{same}(a) := Q_\tau^{Ask}(a)$ (respectively $Q_\tau^{opposite}(a) := Q_\tau^{Bid}(a)$) if it was a sell limit order.

We normalize the quantities by the best opposite to obtain $\rho_\tau(a) = Q_\tau^{same}(a)/Q_\tau^{opposite}(a)$ the fraction of the quantity on the same side of the limit order over the quantity on the opposite side. It is then easy to average over the transactions indexed by timestamps τ to obtain an estimate of this expected ratio for one class of agent:

$$R(a) = \frac{1}{\text{Card}(\mathcal{T})} \sum_{\tau \in \mathcal{T}} \rho_\tau(a), \quad \lim_{\text{Card}(\mathcal{T}) \rightarrow +\infty} R(a) = \mathbb{E}_\tau \left(\frac{Q_\tau^{same}(a)}{Q_\tau^{opposite}(a)} \right). \quad (4)$$

It is even possible to control a potential bias by using the same number of buy and sell executed limit orders to compute this “neutralized” average:

$$R^n(a) = \frac{1}{\text{Card}(\mathcal{T}(\text{buy}))} \sum_{\tau \in \mathcal{T}(\text{buy})} \rho_\tau(a) + \frac{1}{\text{Card}(\mathcal{T}(\text{sell}))} \sum_{\tau \in \mathcal{T}(\text{sell})} \rho_\tau(a). \quad (5)$$

Figure 2.a shows the average state of the imbalance (via some estimates of $R^n(a)$, on AstraZeneca from January 2013 to August 2013) for each class of agent (see Tables 3, 4, 5 for lists of NASDAQ-OMX memberships used to identify agents classes). One can see the state of the imbalance is different for each class given it “accepted” to transact via a limit order:

⁵For each transaction, we have a buyer ID, seller ID, a size, a price and a timestamp.

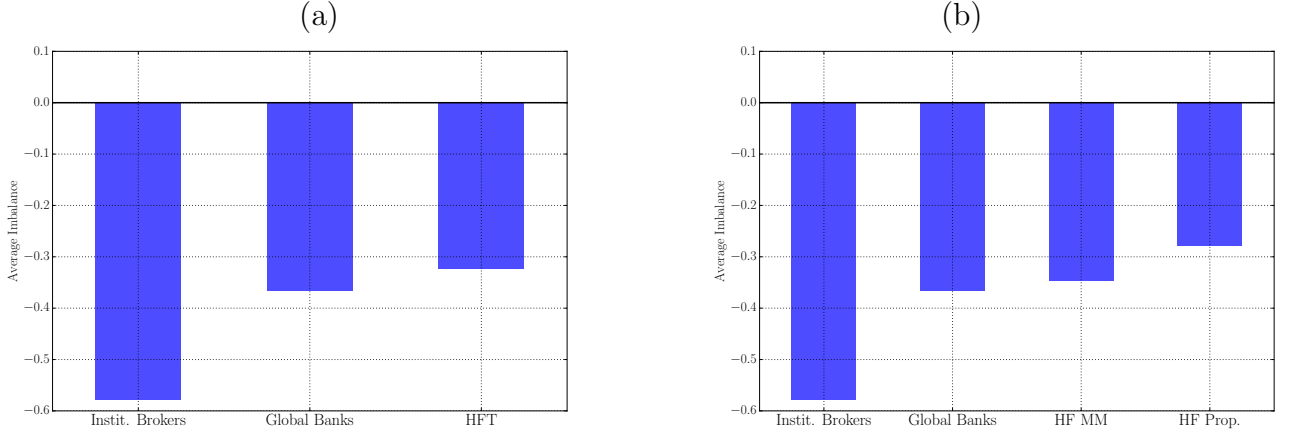


Figure 2: Comparison of neutralized orderbook Imbalance $R^n(a)$ at the time of a trade via a limit order (a) for institutional brokers, global investment banks and High Frequency Participants. (b) Shows a split of HFP between market makers and proprietary traders. It is clear each type of participant has a different strategy. Data are the ones for AstraZeneca (2013-01 to 2013-09).

- Institutional brokers accept a transaction when the imbalance is largely negative, i.e. they buy using a limit order while the price is going down. It generates a large adverse selection: they would have waited a little more, the price would have been cheaper. It may be because they do not pay enough attention to the orderbook, or they make this choice because they have to buy fast from risk management reasons on their clients' orders (in our model we have a waiting cost parameter c that can probably be used to inject such urgency in the model).
- High Frequency Participants (HFP) accept a transaction when the imbalance is around one half of the one when Institutional brokers accept a trade: they make another choice. For sure they look more at the orderbook state before taking a decision. Moreover they could be more opportunistic: ready to wait the perfect moment instead of being led by urgency considerations.

If we split HFP between more market making-oriented ones and proprietary trading ones on Figure 2.b we see

- market makers (probably for urgency reasons: they have to alternate buys and sells), accept to trade when the imbalance is more negative than the average of HFP. They are probably paid back from this adverse selection by bid-ask spread gains (see [Menkveld, 2013]);
- while proprietary traders are from far the most opportunistic participants of our panel, leading them to have a less intense imbalance when they trade via limit orders: they seem to be the ones less suffering from adverse selection.
- Global Investment banks are in between: or their activity is a mix of client execution and proprietary trading (hence we perceive the imbalance when they accept a trade as an average of the two upper ones), or they have specific strategies to accept transactions via limit orders, or they invest a little less than HFP in low latency technology, but more than institutional brokers.

Obviously each class of agent seems to (be able to) exploit differently the state of the orderbook before accepting or not a transaction.

The value of imbalance for market participants. Now we know classes of agents take differently into account the state of orderbook imbalance to accept or not a transaction via a limit order, one can ask what could be the value of such an “*high frequency market timing*”.

We attempt to measure this value with a combination of NASDAQ-OMX labelled transactions and our synchronized market data. That for, we compute the midprice move immediately before and after a class of participant a accepted to transact via a limit order:

$$\Delta P_{\delta t}^{mid}(\tau, a) = \frac{P_{\tau+\delta t}^{mid} - P_{\tau}^{mid}}{\bar{\psi}} \cdot \epsilon_{\tau}(a), \quad (6)$$

where $\epsilon_{\tau}(a)$ is the “sign” of the transaction (i.e. +1 for a buy and -1 for a sell) and $\bar{\psi}$ is the average bid-ask spread on the considered stock.

A “price profile⁶” around a trade is the averaging of this price move as a function of δt (between -5 minutes and +5 minutes); it is an estimate of the “expected price profile” around a trade:

$$p_a(\delta t) = \frac{1}{\text{Card}(\mathcal{T})} \sum_{\tau \in \mathcal{T}} \Delta P_{\delta t}^{mid}(\tau, a), \quad \lim_{\text{Card}(\mathcal{T}) \rightarrow +\infty} p_a(\delta t) = \mathbb{E}_{\tau} \Delta P_{\delta t}^{mid}(\tau, a).$$

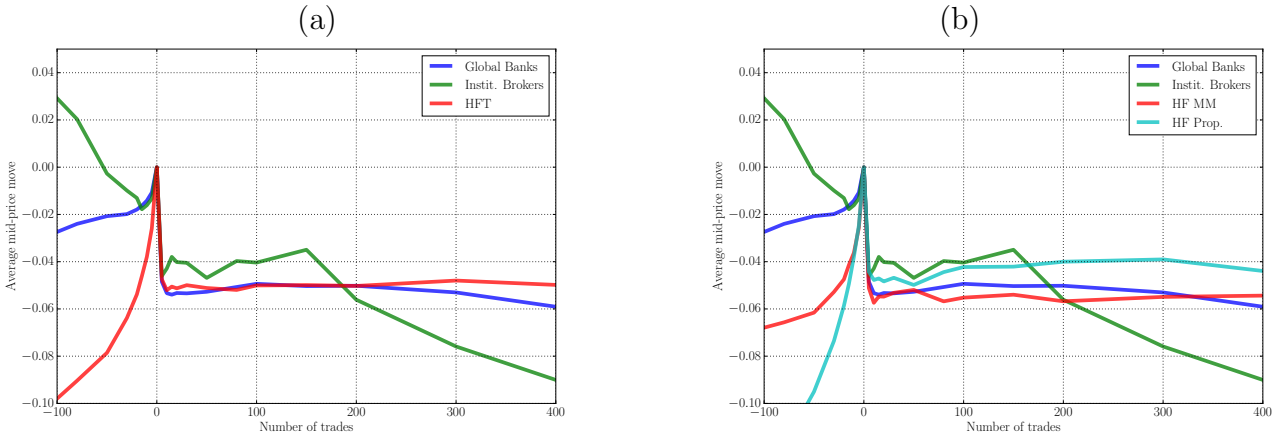


Figure 3: Midprice move relative to its position when a limit order is executed for (a) an High Frequency Market Maker, a regional investment bank, and an institutional broker; (b) makes the difference between HF market makers and HF proprietary traders. AstraZeneca (2013-01 to 2013-09).

Figure 3.a and 3.b show the price profiles of our three classes of participants, exhibiting real differences between them. First of all, it confirms the conclusions we draw from Figure 2. Since it is always interesting to have a look at dynamical measures of liquidity (see [Lehalle, 2014] for a defense of the use of more dynamical measures of liquidity instead of plain averages):

- It is clear Institutional brokers (green line) are buying while the price is going down. Would they have bought later, they would have obtained a cheaper price. As underlined early they probably do it by purpose: they can have urgency reasons or they are using a “trading benchmark” that does pay more attention to peg to the executed volume than to the execution price (see [Lehalle et al., 2013, Chapter 3] for details about brokers’ benchmarks).

⁶Note this “price profiles” are now used as a standard way to study the behaviour of high frequency traders in academic papers, see for instance [Brogaard et al., 2012] or [Biais et al., 2016].

- We can see the difference between High Frequency Participants (HFP) and Global investment banks comes from the price dynamics *before the trade via a limit order*: for Investment banks the price was more or less stable, and had to go down so that the limit order is executed. For HFP the price clearly went up *before they bought with a limit order*. This implies they inserted their limit order shortly before the trade. In our framework we will see how cancelling and reinserting limit orders can be a way to implement an optimal strategy.
- On Figure 3.b we see the difference between HF market makers and HF proprietary traders: the latter succeed in inserting buy (respectively sell) limit orders and obtaining transaction while the price is clearly going up (resp. down). After the trade, one can read a difference between them and HF market makers: proprietary traders suffer from less adverse selection (the cyan curve is a little higher than the red one).

These charts show there is a value in taking liquidity imbalance into account. In the following sections of this paper we will theoretically show how paying attention to orderbook imbalance can be valued to insert or cancel limit orders.

The role of latency. Without a fast enough access to the exchange servers, a participant could know the best action to perform (insert or cancel a limit order), but not be able to implement it before an unexpected transaction. Since low latency has a cost, some participant may decide to ignore this information and do not access to market feeds, orderbooks states, etc.

In the following sections we will not only provide a theoretical framework to “optimally” exploit orderbook dynamics for limit order placement, but also study its sensitivity to latency, showing how latency can destroy the added value of understanding orderbook dynamics.

In our theoretical framework, we will not study the situations in which the participant knows the best action but cannot implement it on time. We will rather reproduce the case of a participant having access to the state of the orderbook at a lower frequency than another. In practical cases it will mean this participant will see the state of liquidity with delay.

4 The Dynamic Programming Principle Applied to Limit Order Placement

4.1 Formalisation of the Model

We will express the control problem for a buy order of a deterministic size \mathbf{q} . It can be changed to a sell order with ease. Since our value function is linear in \mathbf{q} , this deterministic atomic quantity can be replaced by any random variable independent from other variables with no change.

Let \mathbf{q} be a unit limit order inserted in the first Bid limit of the order book. This order is followed by a quantity Q^{After} and preceded by a quantity Q^{Before} . The first opposite limit has a quantity Q^{Opp} . The quantities Q^{After} , Q^{Before} and Q^{Opp} are multiples of the unit quantity \mathbf{q} (see Figure 4 for an illustration).

For simplicity, we neglect the quantity \mathbf{q} and set:

$$Q^{Same} = Q^{Before} + Q^{After}. \quad (7)$$

It could be changed to $Q^{Same} = Q^{Before} + Q^{After} + \mathbf{q}$; such a case can be written and studied similarly to what we propose here. The results would nevertheless be slightly different.

Limit order book dynamics. At the beginning, we don't differentiate between a cancellation order and a market order. The order book dynamic can be modeled by four counting processes (see Figure 4):

- A counting process $N_t^{Opp,+}$ with an intensity $\lambda^{Opp,+}(Q^{Opp}, Q^{Same})$ representing the inserted orders at the opposite limit.
- A counting process $N_t^{Opp,-}$ with an intensity $\lambda^{Opp,-}(Q^{Opp}, Q^{Same})$ representing the canceled orders at the opposite limit.
- A counting process $N_t^{Same,+}$ with an intensity $\lambda^{Same,+}(Q^{Opp}, Q^{Same})$ representing the inserted orders at the Bid (i.e. same) limit.
- A counting process $N_t^{Same,-}$ with an intensity $\lambda^{Same,-}(Q^{Opp}, Q^{Same})$ representing the canceled orders at the Bid limit.

In this model, these four counting processes depend only on the first limit quantities. Moreover, the Bid-Ask symmetry provides us with the following relation:

$$\begin{cases} \lambda^{Opp,+}(Q^{Opp}, Q^{Same}) = \lambda^{Same,+}(Q^{Same}, Q^{Opp}) \\ \lambda^{Opp,-}(Q^{Opp}, Q^{Same}) = \lambda^{Same,-}(Q^{Same}, Q^{Opp}) \end{cases} \quad (8)$$

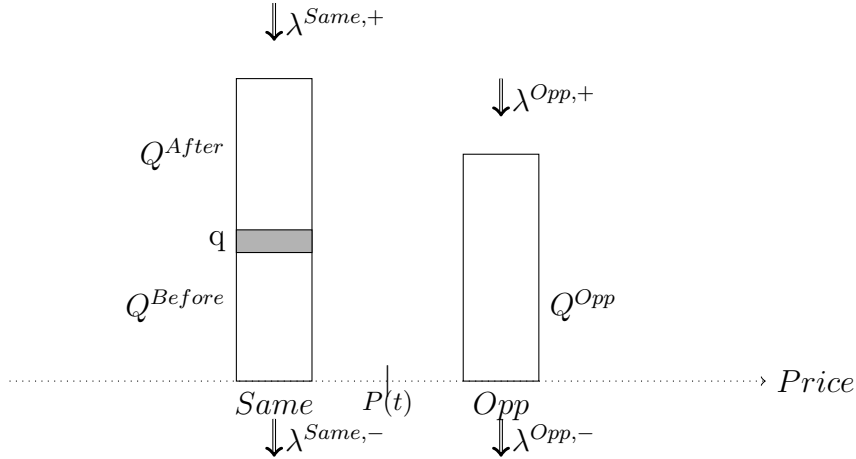


Figure 4: Diagram of flows affecting our orderbook model.

Hence, the size of the first limits can be written as long as none of them is negative :

$$\begin{cases} Q^{Opp}(t+dt) &= Q^{Opp}(t) + dN_t^{Opp,+} - dN_t^{Opp,-} \\ Q^{Before}(t+dt) &= Q^{Before}(t) - dN_t^{Same,-} \\ Q^{After}(t+dt) &= Q^{After}(t) + dN_t^{Same,+} - 1_{Q^{Before} \leq 0} dN_t^{Same,-} \end{cases} \quad (9)$$

What happens when Q^{After} , Q^{Before} or Q^{Opp} is totally consumed : First of all, we neglect the probability that at least two of these three events happen simultaneously.

1. When $Q^{Opp}(t) = 0$. The price increases by one tick (keep in mind for a buy order, the *opposite* is the ask side). Then, we discover a new opposite limit and a new bid quantity

is inserted into the bid-ask spread (on the bid side) by other market participants (see Figure 5). It reads

$$\begin{cases} Q^{Opp}(t + dt) &= Q^{Disc}(Q_t^{Same}, Q_t^{Opp}) \\ Q^{Before}(t + dt) &= Q^{Ins}(Q_t^{Same}, Q_t^{Opp}) \\ Q^{After}(t + dt) &= 0 \end{cases} \quad (10)$$

Q^{Disc} is the “*discovered quantity*” and Q^{Ins} the “*inserted quantity*”. We model them via random variables that can be function of the orderbook state *before the price changing event*.

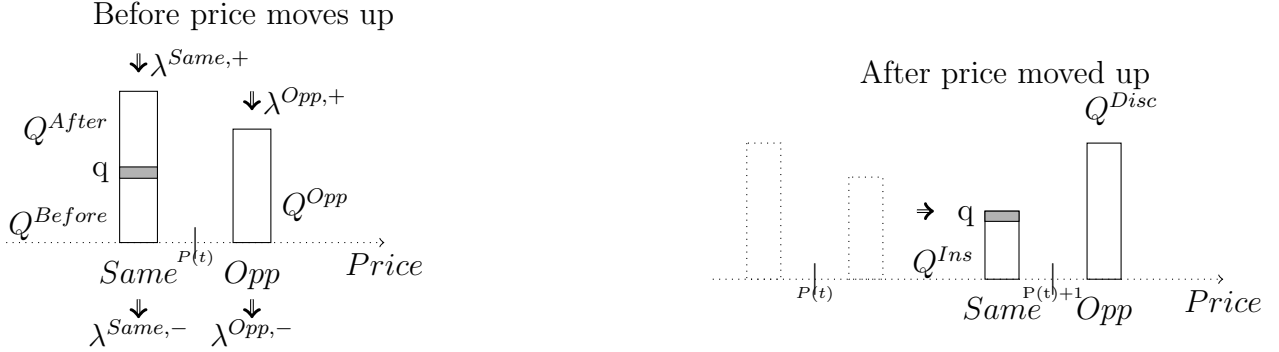


Figure 5: Diagram of an upward price change for our model.

2. When $Q^{Before}(t) = 0$. The optimized limit order is going to be executed. The new state of the limit order book is the following:

$$\begin{cases} Q^{Opp}(t + dt) &= Q^{Opp}(t) + dN_t^{Opp,+} - dN_t^{Opp,-} \\ Q^{Before}(t + dt) &= 0 \\ Q^{After}(t + dt) &= Q^{After}(t) + dN_t^{Same,+} - dN_t^{Same,-} \end{cases} \quad (11)$$

3. When moreover $Q^{After}(t) = 0$ The price decreases by one tick. then, we discover a new quantity on the Bid side and market makers insert a new quantity on the opposite side such as:

$$\begin{cases} Q^{Opp}(t + dt) &= Q^{Ins}(Q_t^{Opp}, Q_t^{Same}) \\ Q^{Before}(t + dt) &= Q^{Disc}(Q_t^{Opp}, Q_t^{Same}) \\ Q^{After}(t + dt) &= 0 \end{cases} \quad (12)$$

If the optimized limit order was in the book: it has been executed. Otherwise the price moves down and the trader has the opportunity to reinsert a limit order on the top of Q^{Disc} (see Figure 6 for a diagram).

The control. We consider two types of control $C = \{s, c\}$:

- c (like *continue*): stay in the order book;
- s (like *stop*): cancel the order and wait for a better orderbook state to reinsert it at the top of Q^{same} (Q^{bid} for our buy order). This control will essentially be used to avoid adverse selection, i.e. avoid to obtain a transaction just before a price decrease.

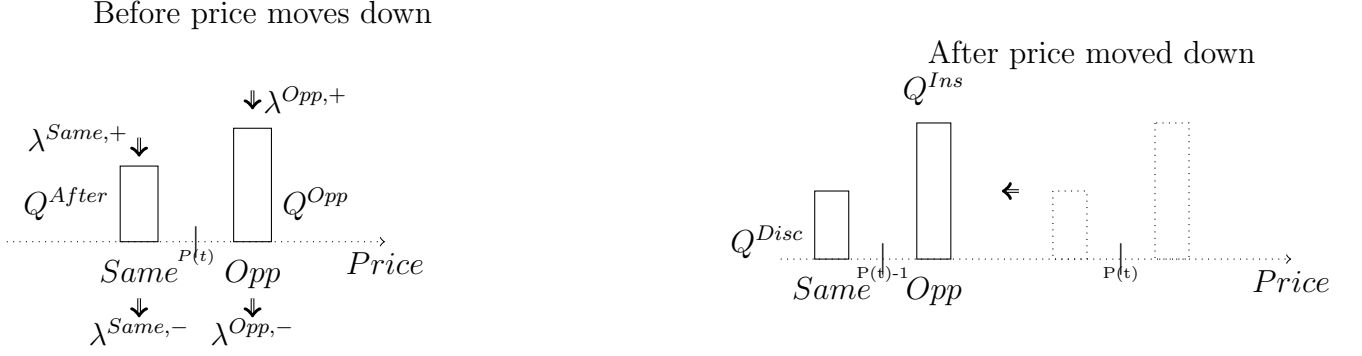


Figure 6: Diagram of an downward price move in our model.

If at the end of T_f periods the order hasn't been executed, we cross the spread to guarantee execution. Once the order is executed, we will compare the execution price to a benchmark price (*microprice*) P_∞ that we will describe further.

We set the time step to Δ_t and the final time to T_f . Let $t_0 = 0 < t_1 \cdots t_{f-1} < t_f = T_f$ different instants at which the order book is observed, such as $t_n = n\Delta_t$ for all $n \in \{0, 1, \dots, f\}$. We add a terminal time T_{f+1} such as $T_{f+1} = (f+1)\Delta_t$.

Under the following assumption: Between two consecutive instants t_n and t_{n+1} for all $n \in \{0, 1, \dots, f-1\}$, only five cases can occur:

- 1 unit quantity is added at the Bid side;
- 1 unit quantity is consumed at the Bid side;
- 1 unit quantity is added at the opposite side;
- 1 unit quantity is consumed at the opposite side;
- nothing happens.

We neglect the situation of at least two cases occurring during the same time interval (the probability of such conjunctions are of the orders of λ^2 , hence our approximation remains valid as far as $\lambda^2 dt$ is small compared to one and λ^2 is small compared to λ).

Framework 1 (Our setup in few words.). *In short, our main assumptions are:*

- *only one limit order of atomic quantity q is controlled, it is small enough to have no influence on orderbook imbalance;*
- *decrease of queue sizes at first limits is caused by transactions only (i.e. no difference between cancellation and trades);*
- *queues decrease or increase by one quantity only;*
- *the intensities of point processes (including the ones driving quantities inserted into the bid-ask spread, and driving the quantity discovered when a second limit becomes a first limit) are functions of the quantities at best limits only;*
- *no notable conjunction of multiple events.*

We introduce the following Markov chain $U_n = (P_n, Q_n^{Before}, Q_n^{After}, Q_n^{Opp}, Exec_n)$ where:

- P_n is the mid-price at time t_n that takes value in \mathbb{R}_+ .

- Q_n^{Before} is the Q^{Before} size at time t_n that takes value in \mathbb{N} .
- Q_n^{After} is the Q^{After} size at time t_n that takes value in \mathbb{N} .
- Q_n^{Opp} is the Q^{Opp} size at time t_n that takes value in \mathbb{N} .
- $Exec_n$ is an additional variable that takes value in $\{-1, 0, 1\}$. $Exec_n$ equals to 1 when the order is executed at time t_n , 0 when the order is not executed at time t_n and -1 (a “cemetery state”) when the order has been already executed before t_n . Initially, we fix $Exec_0 = 0$.

In the same way, we define $N_n^{Same,+}$, $N_n^{Same,-}$, $N_n^{Opp,+}$ and $N_n^{Opp,-}$ as the values of the counting processes $N_t^{Same,+}$, $N_t^{Same,-}$, $N_t^{Opp,+}$ et $N_t^{Opp,-}$ at time t_n . The transition probabilities of the markov chain U_n are detailed in Appendix A.

The terminal constraint. The microprice $P_{\infty,k}$ is defined such as:

$$P_{\infty,k} = \mathbb{E}(P_{\infty}|\mathcal{F}_k) = F(Q_k^{Opp}, Q_k^{Same}, P_k) = P_k + \frac{\alpha}{2} \cdot \frac{Q_k^{Same} - Q_k^{Opp}}{Q_k^{Opp} + Q_k^{Same}} \quad \forall k \in \{0, 1, \dots, f\}.$$

Where \mathcal{F}_k is the filtration associated to U_k such as $\mathcal{F}_k = \sigma(U_n, n \leq k)$ and α is a parameter that represents the sensitivity of futur prices to the imbalance.

The execution price $P_{Exec,k}$ is defined $\forall k \in \{0, 1, \dots, f\}$ such as :

$$P_{Exec,k} = \begin{cases} P_k - \frac{1}{2} & \text{when } Exec_k = 0 \\ P_k + \frac{1}{2} & \text{when } Exec_k \in \{-1, 1\} \text{ (when a market order is set).} \end{cases}$$

Let k_0 be the execution time: $k_0 = \inf(k \geq 0, Exec_k = 1) \wedge f$. Then, the terminal valuation can be written:

$$Z_{k_0} = P_{\infty, (k_0+1)} - P_{Exec, (k_0)}$$

Let \mathcal{U} the set of all progressively measurable processes $\mu := \{\mu_k, k < f\}$ valued in $\{s, c\}$. This problem can be written as a stochastic control problem :

$$V_{U_0, f} = \sup_{\mu \in \mathcal{U}^{U_0, \mu}} \mathbb{E} \left(\sum_{i=1}^{f-1} g_i(U_i, \mu_i) + g_f(U_f) \right)$$

Where $g_i(U_i, \mu_i) = Z_i$ when $Exec_i = 1$ and $\mu_i = c$ and 0 otherwise for all $i \in \{1, \dots, f-1\}$, and $g_f(U_f) = Z_f$ when $Exec_f \in \{0, 1\}$ and 0 otherwise.

In other words we want to maximize $V_{U_0, f} = \sup_{\mu \in \mathcal{U}^{U_0, \mu}} \mathbb{E}(Z_{k_0})$ which can be reached by the dynamic programming algorithm:

$$\begin{cases} G_f = Z_f \\ G_n = \max(P_n^c G_{n+1}, P_n^s G_{n+1}) \quad \forall n \in \{0, 1, \dots, f-1\} \end{cases} \quad (13)$$

Where P_n represents the transition matrix of the markov chain U_n .

5 A Qualitative Understanding

The equation (13) provides an explicit forward-backward algorithm that can be solved numerically. The following part present the simulation results. For more details about the forward-backward algorithm see Appendix A. This section comments simulation results.

We are going to compare two situations:

- The first one called **(NC)** corresponds to the case when no control is adopted (i.e we always stay in the orderbook).
- The second one called **(OC)** corresponds to the optimal control case when both controls "c" and "s" are considered.

Moreover, our simulation results are given for two different cases :

- **Framework (CONST):** The intensities of insertion and cancellation are constant: $\lambda_k^{Same,+} = \lambda_k^{Opp,+} = 0.06$ and $\lambda_k^{Same,-} = \lambda_k^{Opp,-} = 0.1 \ \forall k \in \{0, 1, \dots, f\}$. Under **(CONST)**, the inserted quantities Q^{Ins} and discovered quantities Q^{Disc} are constant too.
- **Framework (IMB):** The intensities of cancellation and insertion are functions of the imbalance such as $\forall k \in \{0, 1, \dots, f\}$:

$$\begin{aligned} \lambda_k^{Opp,+} \left(Q_k^{Opp}, Q_k^{Same} \right) &= \lambda_k^{Same,-} \left(Q_k^{Opp}, Q_k^{Same} \right) = \frac{Q_k^{Opp}}{\beta \cdot (Q_k^{Opp} + Q_k^{Same})} \\ \lambda_k^{Opp,-} \left(Q_k^{Opp}, Q_k^{Same} \right) &= \lambda_k^{Same,+} \left(Q_k^{Opp}, Q_k^{Same} \right) = \frac{Q_k^{Same}}{\beta \cdot (Q_k^{Opp} + Q_k^{Same})} \end{aligned}$$

Where $\frac{1}{\beta}$ is a predictability parameter that represents the sensitivity of order flows to the imbalance.

Moreover, under **(IMB)**, inserted and discovered quantities are computed the following way:

- **When Q_k^{Opp} is totally consumed**, we set $Q_k^{Disc} = \lfloor \theta_{disc} \cdot Q_k^{Opp} \rfloor$ and $Q^{Ins} = \lfloor \theta_{ins} \cdot Q_k^{Same} \rfloor$. Where θ_{disc} and θ_{ins} are coefficients associated to liquidity and $\lfloor \cdot \rfloor$ is the upper rounding .
- Similarly **when Q^{Same} is totally consumed**, we set $Q^{Disc} = \lfloor \theta_{disc} \cdot Q^{Same} \rfloor$ and $Q^{Ins} = \lfloor \theta_{ins} \cdot Q^{Opp} \rfloor$.

This kind of relations is compatible with empirical findings of [Huang et al., 2015b] and different from [Cont and De Larrard, 2013] in which $Q^{disc} = Q^{ins}$ is independant of liquidity imbalance.

5.1 Numerically Solving the Control Problem

5.1.1 Anticipation of Adverse Selection

The cancellation is used by the optimal strategy to avoid adversion selection. For instance, when the quantity on the same side is extremely lower than the one on the opposite side, it is expected to cancel the order to wait for a better future opportunity. The optimal control takes in consideration this effect and cancels the order when such a high adverse selection effect is present.

We keep notations of Section 4. Let $\mu := \{\mu_k, k < f\}$ a control, we define $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec}) = \mathbb{E}_{U_0, \mu}(P_{\infty, (k_0+1)} - P_{\text{Exec}, (k_0)})$. $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec})$ depends on the control μ , the initial state of the orderbook U_0 and the terminal time f . The dynamics of the quantity $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec})$ can be written :

$$\begin{aligned} \mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec}) &= \sum_{i=1}^{f-1} \sum_{\substack{\text{states } U_i \\ \text{under } \mu}} 1_{\text{Exec}_i=1} \cdot (P_{\infty, (i+1)} - P_i) \cdot p_{U_i} \\ &+ \sum_{\substack{\text{states } U_f \\ \text{under } \mu}} 1_{\text{Exec}_f \in \{0,1\}} \cdot (P_{\infty, (f+1)} - P_f) \cdot p_{U_f} \end{aligned}$$

Where $p_{U_i} \forall i \in \{0, 1, \dots, f\}$ represents the probability to reach the states U_i starting from the initial point U_0 under μ . The quantity p_{U_i} can easily be computed knowing the transition matrix of the markov chain U_n (cf. Appendix A). Thusly, thanks to the former equation, the quantity $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec})$ can be directly computed by a forward algorithm that visits all the possible states of the markov chain U_n under the control μ .

The quantity $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec})$ is interesting because it corresponds exactly to the quantity to be maximize by our optimal control problem and represents as well the profitability/trade of an agent.

Let μ^c the control associated to the case where we always choose to stay in the limit orderbook (i.e NC). The Figure 7.a represents the variation of $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ (i.e optimal strategy OC) and $\mathbb{E}_{U_0, \mu^c}(\Delta P | \text{Exec})$ (i.e NC) when the initial imbalance of the orderbook moves under (CONST). In Figure 7.a, Blue points refer to the cases where it is better to stay in the orderbook at the beginning while red points refer to the cases where it is better to cancel the order at $t = 0$. The initial parameters are fixed such as $\lambda^{Same, +} = \lambda^{Opp, +} = 0.06$, $\lambda^{Same, -} = \lambda^{Opp, -} = 0.1$, $\alpha = 24$, $Q^{Disc} = 4$, $Q^{Ins} = 2$, $f = 4$ and $P_0 = 10$. Moreover, the different values of the initial imbalance are obtained by varying Q_0^{Opp} from 2 to 10 and Q_0^{After} from 2 to 10 while Q_0^{Before} is kept constant equal to 0. We keep in mind that the order is executed when Q^{After} is consumed (cf. appendix A for more details).

In Figure 7.b, we represent the same thing as in Figure 7.a but under the framework (IMB). In Figure 7.b, The initial parameters are fixed such as $\beta = 3$, $\theta_{disc} = 3$, $\theta_{ins} = 0.6$, $\alpha = 24$, $f = 4$ and $P_0 = 10$. Similarly, the different values of the initial imbalance are obtained by varying Q_0^{Opp} from 2 to 10 and Q_0^{After} from 2 to 10 while Q_0^{Before} is kept constant equal to 0.

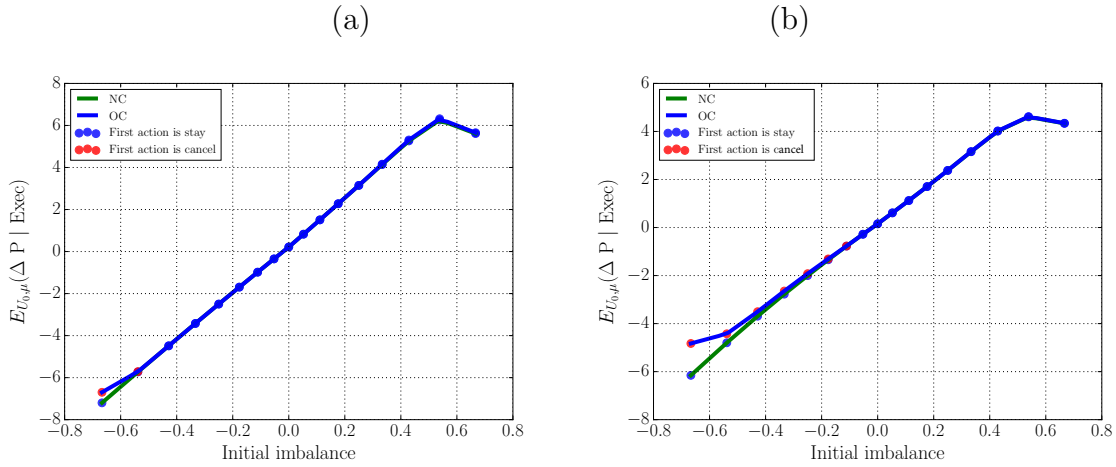


Figure 7: $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ move relative to initial imbalance when intensities are constant (CONST) in (a) and when they are variable (IMB) with $\beta = 3$ in (b).

In Figure 7, the expected price move given an execution has been obtain according to frameworks (CONST) (Figure 7.a) and (IMB) (Figure 7.b), comparing the simulated optimal strategy (OC) to the simple “join the bid” (NC) one.

The **main effect** to note on these curves is the way the optimal control anticipate adversion selection. When Imbalance is highly negative, we cancel first the order (red points) to take advantage from a better futur opportunity. However, as our order has a unit size, the algorithm wait until the last moment to cancel the order. That’s why, we took $\alpha = 24$ to accentuate the imbalance effect. We notice that the (OC) case provides better result than the (NC) case (cf. Figure 8). This point is going to be detailed further.

Appendix C explains the downward slopes at the left of Figures 7.a and 7.b.

5.1.2 Price Improvement comes from avoiding adverse selection

As expected, the results obtained in the optimal control (OC) case are better than the ones in the non-controlled (NC) case : by cancelling and taking into account liquidity imbalance, one can be more efficient than just staying in the orderbook.

Figure 8.a represents the variation of the price improvement due to the control: $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec}) - \mathbb{E}_{U_0, \mu^c}(\Delta P | \text{Exec})$ when the initial imbalance of the orderbook moves, under (CONST) and (IMB).

Similarly, The Figure 8.b represents the variation of $\mathbb{E}_{U_0, \mu^*}(P | \text{Exec}) - \mathbb{E}_{U_0, \mu^c}(P | \text{Exec})$ when the initial imbalance of the limit orderbook moves, under (CONST) and (IMB).

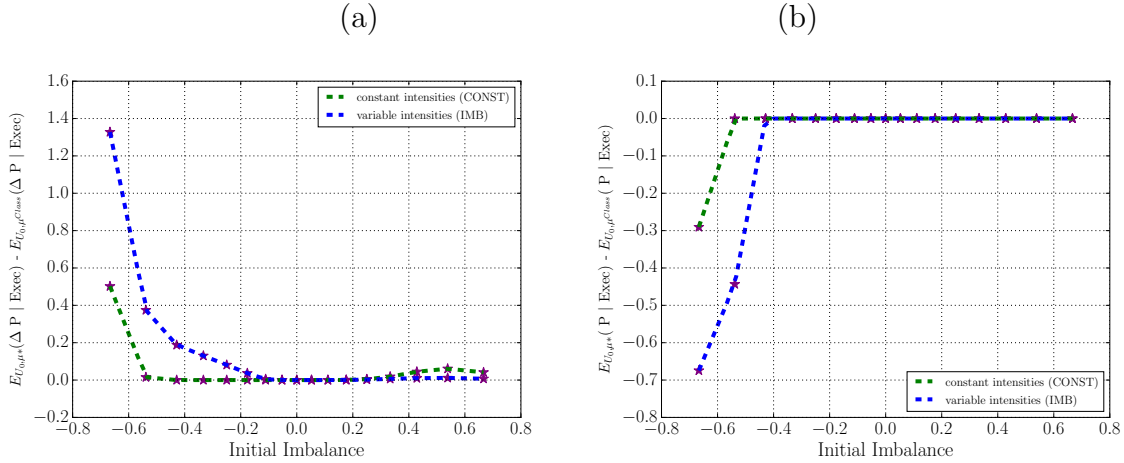


Figure 8: (a) $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec}) - \mathbb{E}_{U_0, \mu^c}(\Delta P | \text{Exec})$ move relative to initial imbalance under (CONST) and (IMB). (b) $\mathbb{E}_{U_0, \mu^*}(P | \text{Exec}) - \mathbb{E}_{U_0, \mu^c}(P | \text{Exec})$ move relative to initial imbalance under (CONST) and (IMB).

Figure 8 deserves the following comments. As expected, the optimal control provides better results than a blind “follow the bid” strategy. In Figure 8.a we can see the price improvement is non-negative because our algorithm maximize $\mathbb{E}(\Delta P)$. When the imbalance is highly positive, the difference is close to 0 while when imbalance is highly negative the advantage of acting optimally becomes higher than 0 by avoiding adverse selection. Similarly, in Figure 8.b we can see that the optimal strategy allows to buy with a lower average price when imbalance is highly negative by preventing from adverse selection. Moreover, this effect can be accentuated when intensities depend on the imbalance. ⁷

5.1.3 Average Duration of Optimal Strategies

In brief, the optimal strategy aims to obtain an execution in the best conditions (i.e. With a low adverse selection risk). It can be read on the average lifetime (i.e. “duration”) of the strategy. Figure 9.a compares the average strategy duration in (NC) and (OC) case under the (CONST) framework. The initial parameters are fixed such as $\lambda^{Same, +} = \lambda^{Opp, +} = 0.06$, $\lambda^{Same, -} = \lambda^{Opp, -} = 0.1$, $\alpha = 10$, $Q^{Disc} = 4$, $Q^{Ins} = 2$, $f = 4$ and $P_0 = 10$. Figure 9.b represents the same as Figure 9.a but under framework (IMB). In this case, the initial parameters are fixed such as $\beta = 4$, $\theta_{disc} = 3$, $\theta_{ins} = 0.6$, $\alpha = 10$, $f = 4$ and $P_0 = 10$. Finally, Figure 9.c represents the “stay ratio” (i.e. the proportions of trajectories for which the optimal strategy chooses to not cancel its limit order) in (NC) and (OC) case under (IMB) with respect to the initial imbalance. Figure 9 is computed with the same initial parameters as Figure 9.b. In Figure 9.a,

⁷Indirectly, maximizing $\mathbb{E}_{U_0, \mu}(\Delta P | \text{Exec})$ leads to the minimization of the price.

9.b and 9.c, the different values of the initial imbalance are obtained by varying Q_0^{Opp} from 2 to 10 and Q_0^{After} from 2 to 10 while Q_0^{Before} is kept constant equal to 0.

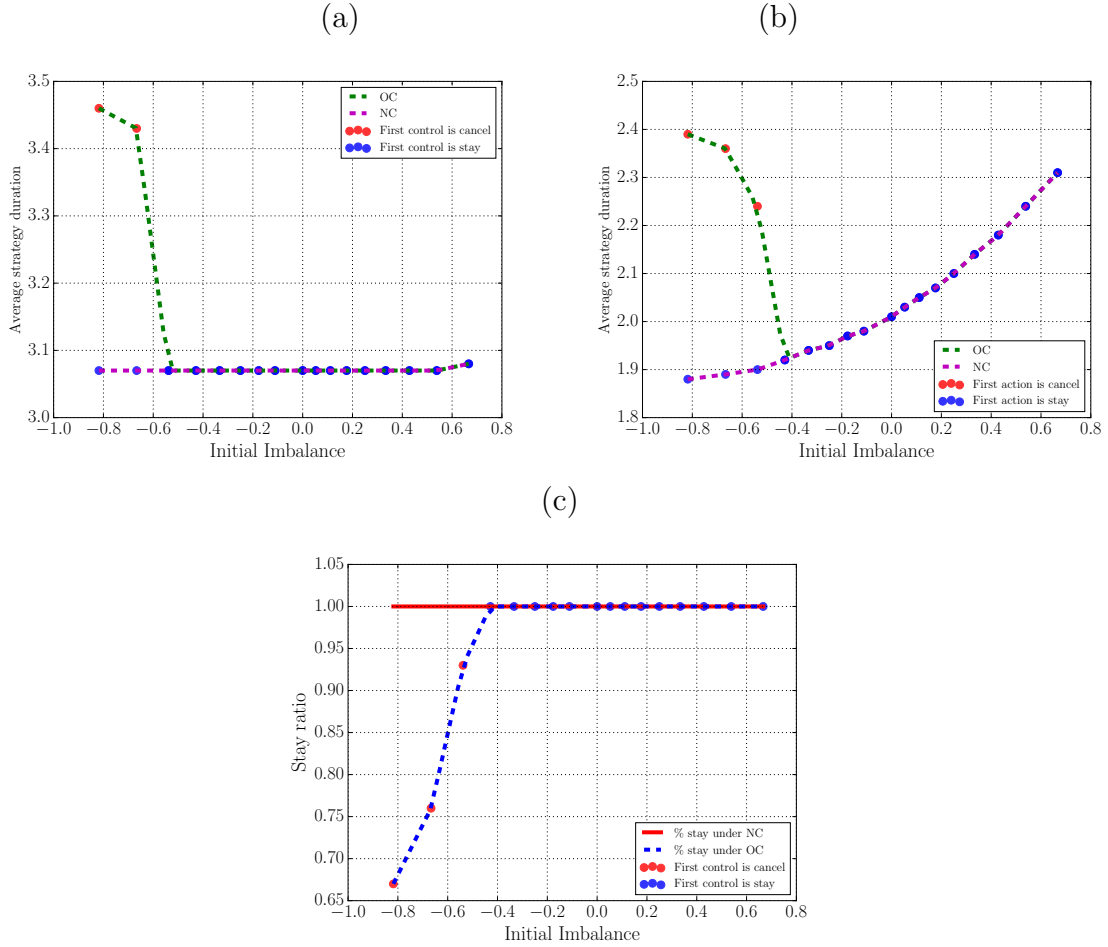


Figure 9: Average strategy duration as a function of (a) the initial imbalance under (CONST) and (b) initial imbalance under (IMB). (c) Stay ratio as a function of the initial imbalance under (IMB).

Figure 9 shows

- In both Figure 9.a and 9.b, the average strategy duration of the optimal control is always higher than the non-optimal one. It is an expected result because the optimal control can cancel its order and hence choose to postpone its execution. Moreover, we can see that the algorithm cancels the order when high adverse selection is present (i.e the imbalance is highly negative under IMB⁸). Consequently, the average strategy duration of the optimal control is strictly greater than the non-optimal one when the imbalance is highly negative.
- In Figure 9.b, when intensities depend on the imbalance (IMB), we can see a decreasing trend of the average strategy duration. In fact, under (IMB), when imbalance is highly positive, more weights are given to events delaying execution. For instance, when imbalance is highly positive, the bid queue is a way larger than the opposite one, then the probability to obtain an execution on the bid side is low : that's why it is expected to have to wait more. Moreover, Figure 9.c shows that we become more active when high adverse selection is present. Actually, when high adverse selection (i.e. negative imbalance) is present, the "stay ratio" decreases and consequently the "cancel ratio" increases.

⁸ close to $t = 0$, the optimal strategy is free to cancel its limit order; but when T_f is close, it has to think about the cost of having to cross the spread in few steps.

5.1.4 Influence of the Terminal Constraint

In this paragraph, we want to shed light on two stylized facts:

1. the optimal strategy performs better under good initial market condition if it has more time left.
2. the optimal strategy becomes highly active close to the terminal time.

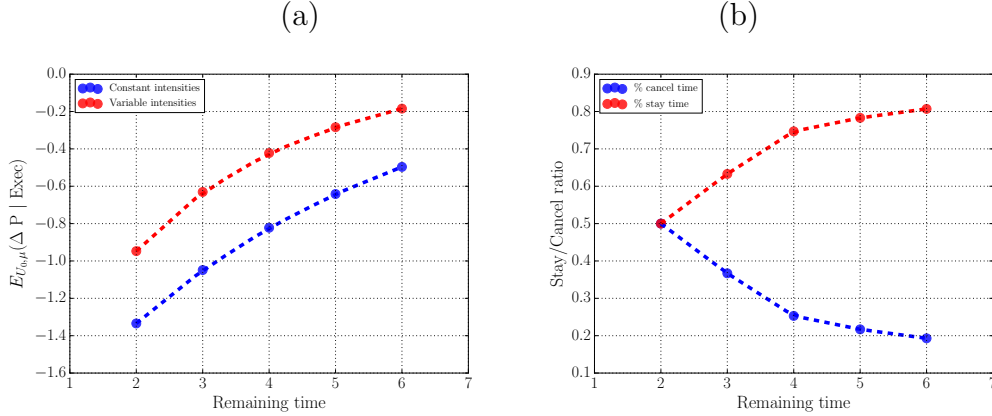


Figure 10: (a) $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ move relative to remaining time under (CONST) and (IMB). (b) stay/cancel ratio move relative to remaining time to maturity under (IMB).

Figure 10.a compares the variation of the adverse selection $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ as a function of the remaining time under (CONST) and (IMB). The initial imbalance is fixed equal to -0.5. Thanks to the Figure 10.a, we can see that the more remaining time, the better for the optimal strategy. However, the concavity of the curve shows that the marginal performance $\frac{\partial \mathbb{E}(\Delta P)_t}{\partial t}$ is decreasing. Moreover, Figure 10.a shows also that $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ may converge to a limit value when maturity time tends to infinity. Since the markov chain U_n is ergodic (cf. [Huang et al., 2015b]), we believe that this limit value is unique and independent of the initial state of the orderbook.

In Figure 10.b, we represent the percentage of times where the optimal strategy cancels its order and the percentage of times where it decides to stay in the orderbook as a function of remaining time under (CONST) and (IMB). The initial imbalance is fixed to -0.5. Thanks to Figure 10.b, we can see that it is optimal to be more active close to $t = T_f$.

5.2 The Price of Latency

In Section 4, we have defined the markov chain U_n which corresponds to a market participant enabled to change his control at each period. A slower participant will not react at each limit orderbook move. Hence, he can be modelled by the markov chain $U_{\tau n}$ where τ corresponds to a latency factor such as $\tau \in \mathbb{N}^*$.

Using notations of the previous sections, we define $Z_{\tau, f}$ as the final constraint associated to the Markov chain $U_{\tau n}$. Thus, we define the latency cost of a participant with a latency factor τ such as :

$$\text{Latency}_{U_0, f}(\tau) = V_{U_0, f} - V_{U_0, f, \tau} \quad \forall \tau \in \mathbb{N}^* \quad (14)$$

Where $V_{U_0, f, \tau} = \sup_{\mu \in \mathcal{U}} \mathbb{E}_{U_0, \mu}(Z_{\tau, f})$.

By adapting the same numerical forward-backward algorithm, the cost of latency can be computed numerically. This cost can be converted into a value: it is the value a trader should accept to pay in technology since he will be rewarded in term of performance.

In Figure 11.a, we can see the variation of the latency cost with respect to the latency factor τ under (CONST) and (IMB). The initial parameters are fixed such as $\beta = 3$, $\theta_1 = 3$, $\theta_2 = 0.6$, $\alpha = 2$, $f = 5$ and $P_0 = 10$. The initial imbalance is fixed to 0.5 with an initial state $Q_0^{After} = 2, Q_0^{Before} = 0$ and $Q_0^{Opp} = 6$.

The Figure 11.b represents the variation of the latency cost for different value of the predictability parameter α when intensities depend on the imbalance.

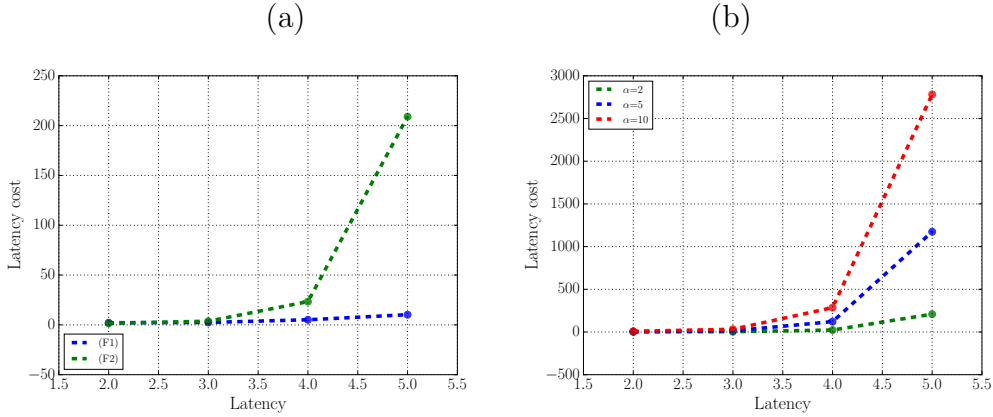


Figure 11: (a) Latency costs as a function of the latency factor τ under (CONST) and (IMB). (b) Latency costs as a function of the latency factor τ under (IMB) for different values of α .

The numerical results shows :

- The latency costs increase exponentially with the latency τ (cf. Figure 11.a). The latency costs are amplified when intensities are function of imbalance (cf. Figure 11.a).
- The latency costs are higher when sensitivity to adverse selection increases (i.e α is big) (cf. Figure 11.b).

Consequently, the added value of exploiting a knowledge on liquidity imbalance is eroded by latency: being able to predict future liquidity consuming flows is of less use if you can't cancel and reinsert your limit orders at each change of the orderbook state. For instance, when two agents act optimally according the same criterion, the faster will have more profits than the slower.

6 Conclusion

We have used NASDAQ-OMX labelled data to show market participants accept or refuse transactions via limit orders as a function of liquidity imbalance. It is not an exhaustive study on this exchange from the north of Europe (we focus on AstraZeneca from January 2013 to September 2013). We first show orderbook imbalance has a predictive power on future mid price move. We then focus on three types of market participants: Institutional brokers, Global Investment Banks (GIB) and High Frequency Participants (HFP). Data show the former accept to trade when the imbalance is more negative (i.e. they buy or sell while the pressure is upward

or downwards) than GIB, themselves accepting a less negative imbalance than HFP. Moreover, when we split HFP between high frequency market makers and high frequency proprietary traders, we see HFPT achieve to buy via limit orders when the imbalance is very small. We complete this analysis with the dynamics of prices around limit orders execution, showing how strategically participants use their limit orders.

Then we propose a theoretical framework to control limit orders when liquidity imbalance can be used to predict future price move. Our framework includes potential adverse selection via a parameter α . We use the dynamic programming principle to provide a way to solve it numerically and exhibit simulations. We show the solutions of our framework have commonalities with our empirical findings.

In a last Section we show how the capability of exploiting imbalance predictability using optimal control decreases with latency: the trader has less time to put in place sophisticated strategies, hence he cannot take profit of any strategy gain.

The difficult point of using limit orders is adverse selection: when you buy, it is easy to obtain a transaction when the price is going down, but it is not a good way to obtain a trade since few seconds later, the price will be better. Nevertheless it is not enough to cancel your order if you know the price will go down (you can rely on liquidity imbalance to have some views on the future price direction), because when you will reinsert it will be on the top of the bid queue, and you will probably never obtain a transaction: the price will go up again before your limit order has been executed.

Our framework includes all these effects, hence our optimal strategy makes the choice between waiting in the queue or leaving it when the probability the price will go down is too high. Of course the position of the limit order in the queue is taken into account by our controller.

This leads to a quantitative way to understand market making and latency: if a market maker is fast enough, he will be able to play this insert, cancel and re-insert game to react to his observation of liquidity imbalance. In our framework we use the difference between the sizes of the first bid and ask queues as a proxy of liquidity imbalance, in the real word market participants can use a lot of other information (like liquidity imbalance on correlated instruments, or realtime news feeds).

In such a context speed can be seen as a protection to adverse selection, potentially reducing the cost to make the market. Within this viewpoint, high frequency actions do not add noise to the price formation process (as opposite to the viewpoint of [Budish et al., 2015]) but allows market makers to offer better quotes. At this stage, we do not conclude speed is good for liquidity because:

- we only focussed on one limit order, we should go towards a framework similar to the one of [Guéant et al., 2013] to conclude on the added value of imbalance for the whole market making process, but it will be too sophisticated at this stage.
- it is not fair to draw conclusions from a knowledge of the theoretical optimal behaviour of one market participant; to go further we should model the game played by all participants, similarly to what have been done in [Lachapelle et al., 2016]. Again it is a very sophisticated work. Moreover, this paper is a first step. We are convinced it is possible to obtain partially explicit formula, to enable more systematic explorations of the influence of parameters (currently our simulations are highly memory consuming). It should allow to confront our results to observed behaviors with accuracy (especially using observed values for our parameters $\alpha, \beta, Q^{Disc}, Q^{Ins}$ and λs).
- Last but not least, any conclusion on the added value of low latency and high frequency market making should take into account market conditions. Its value could change with the level of stress of the price formation.

This work shows imbalance is used by participants, and provides a theoretical framework to play with limit order placement. It can be used by practitioners. More importantly, we hope other researchers will extend our work in different directions to answer to more questions, and we will ourselves continue to work further to understand better liquidity formation at the smallest time scales thanks to this new framework.

Acknowledgements. Authors would like to thank Sasha Stoïkov and Jean-Philippe Bouchaud for discussions about orderbook dynamics and optimal placement of limit orders that motivated this paper. Moreover, authors would like to underline the work of Gary Sounigo (during his Masters Thesis) and Felix Patzelt (during a post-doctoral research), who worked hard at Capital Fund Management (CFM) to understand how to align NASDAQ-OMX labeled transactions to direct datafeed of orderbook records. Long lasting discussions about limit order placement with Capital Fund Management execution researchers (especially with Bence Toth and Mihail Vladkov) influenced very positively our work.

References

- [Almgren and Lorenz, 2006] Almgren, R. and Lorenz, J. (2006). Bayesian adaptive trading with a daily cycle. *Journal of Trading*.
- [Bacry et al., 2014] Bacry, E., Iuga, A., Lasnier, M., and Lehalle, C.-A. (2014). Market Impacts and the Life Cycle of Investors Orders. *Social Science Research Network Working Paper Series*.
- [Besson et al., 2016] Besson, P., Pelin, S., and Lasnier, M. (2016). To cross or not to cross the spread: that is the question! Technical report, Kepler Cheuvreux, Paris.
- [Biais et al., 2016] Biais, B., Declerck, F., and Moinas, S. (2016). Who supplies liquidity, how and when? Technical report, BIS Working Paper.
- [Bouchaud et al., 2004] Bouchaud, J.-P., Gefen, Y., Potters, M., and Wyart, M. (2004). Fluctuations and response in financial markets: the subtle nature of 'random' price changes. *Quantitative Finance*, 4(2):176–190.
- [Brogaard et al., 2012] Brogaard, J., Baron, M., and Kirilenko, A. (2012). The Trading Profits of High Frequency Traders. In *Market Microstructure: Confronting Many Viewpoints*.
- [Budish et al., 2015] Budish, E., Cramton, P., and Shim, J. (2015). The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. *Quarterly Journal of Economics*, 130(4):1547–1621.
- [Cont, 2001] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.
- [Cont and De Larrard, 2013] Cont, R. and De Larrard, A. (2013). Price Dynamics in a Markovian Limit Order Book Market. *SIAM Journal for Financial Mathematics*, 4(1):1–25.
- [Dayri and Rosenbaum, 2015] Dayri, K. and Rosenbaum, M. (2015). Large Tick Assets: Implicit Spread and Optimal Tick Size. *Market Microstructure and Liquidity*, 01(01):1550003.
- [Fodra and Pham, 2013] Fodra, P. and Pham, H. (2013). Semi Markov model for market microstructure.

- [Fricke and Gerig, 2014] Fricke, D. and Gerig, A. (2014). Too Fast or Too Slow? Determining the Optimal Speed of Financial Markets. Technical report, SEC.
- [Guéant et al., 2012] Guéant, O., Lehalle, C.-A., and Fernandez-Tapia, J. (2012). Optimal Portfolio Liquidation with Limit Orders. *SIAM Journal on Financial Mathematics*, 13(1):740–764.
- [Guéant et al., 2013] Guéant, O., Lehalle, C.-A., and Fernandez-Tapia, J. (2013). Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, 4(7):477–507.
- [Harris, 2013] Harris, L. (2013). Maker-taker pricing effects on market quotations. *Unpublished working paper. University of Southern California, San Diego, CA.*
- [Huang et al., 2015a] Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015a). How to predict the consequences of a tick value change? Evidence from the Tokyo Stock Exchange pilot program.
- [Huang et al., 2015b] Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2015b). Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 10(509).
- [Jaisson, 2014] Jaisson, T. (2014). Market impact as anticipation of the order flow imbalance.
- [Kyle, 1985] Kyle, A. P. (1985). Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335.
- [Lachapelle et al., 2016] Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., and Lions, P.-L. (2016). Efficiency of the Price Formation Process in Presence of High Frequency Participants: a Mean Field Game analysis. *Mathematics and Financial Economics*, 10(3):223–262.
- [Lehalle, 2014] Lehalle, C.-A. (2014). Towards dynamic measures of liquidity. *AMF Scientific Advisory Board Review*, 1(1):55–62.
- [Lehalle et al., 2013] Lehalle, C.-A., Laruelle, S., Burgot, R., Pelin, S., and Lasnier, M. (2013). *Market Microstructure in Practice*. World Scientific publishing.
- [Lipton et al., 2013] Lipton, A., Pesavento, U., and Sotiropoulos, M. G. (2013). Trade arrival dynamics and quote imbalance in a limit order book.
- [Lo et al., 2002] Lo, A. W., MacKinlay, C., and Zhan, J. (2002). Econometric models of limit-order executions. *Journal of Financial Economics*, 65(1):31–71.
- [Menkveld, 2013] Menkveld, A. J. (2013). High Frequency Trading and The New-Market Makers. *Journal of Financial Markets*, 16(4):712–740.
- [Moallemi, 2014] Moallemi, C. C. (2014). The Value of Queue Position in a Limit Order Book. In Abergel, F., Bouchaud, J.-P., Foucault, T., Lehalle, C.-A., and Rosenbaum, M., editors, *Market Microstructure: Confronting Many Viewpoints*. Louis Bachelier Institute.
- [Stoïkov, 2014] Stoïkov, S. (2014). Time is Money: Estimating the Cost of Latency in Trading. In Abergel, F., Bouchaud, J.-P., Foucault, T., Lehalle, C.-A., and Rosenbaum, M., editors, *Market Microstructure: Confronting Many Viewpoints*, Paris. Louis Bachelier Institute. sasha14book.

- [Tóth et al., 2012] Tóth, B., Eisler, Z., Lillo, F., Kockelkoren, J., Bouchaud, J. P., and Farmer, J. D. (2012). How does the market react to your order flow? *Quantitative Finance*, 12(7):1015–1024.
- [van Kervel and Menkveld, 2014] van Kervel, V. and Menkveld, A. (2014). Do High-Frequency Traders Engage in Predatory Trading? Technical report, VU University Amsterdam.
- [Zarinelli et al., 2015] Zarinelli, E., Treccani, M., Farmer, J. D., and Lillo, F. (2015). Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, 1(2). lastlillo.

A Transition probabilities of the markov chain U_n

When the first limits are totally consumed, new quantities Q_n^{Disc} and Q_n^{Ins} are inserted in the order book. We introduce then μ_n the joint distribution of the random variables Q_n^{Disc} et Q_n^{Ins} at time t_n . We assume these two variables are independent from their past and independent from the counting processes $N^{Same,+}$, $N^{Same,-}$, $N^{Opp,-}$ and $N^{Opp,+}$. However, Q_n^{Disc} and Q_n^{Ins} can be correlated at time t_n .

Let $n \in \{0, 1, \dots, f\}$, $p \in \mathbb{R}_+$, $q^{bef} \in \mathbb{N}$, $q^{aft} \in \mathbb{N}$, $q^{opp} \in \mathbb{N}$, $q^{disc} \in \mathbb{N}$, $q^{ins} \in \mathbb{N}$ and $e \in \{-1, 0, 1\}$

- **When the order has been executed before t_n (i.e $e = 1$ or $e = -1$), then :**

$$\mathbb{P}^{(c)}(U_{n+1} = (p, q^{bef}, q^{aft}, q^{opp}, -1) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) = 1 \quad (15)$$

When the order is executed a dead center is reached and both quantities and the price remain unchanged.

- **When the order isn't executed at t_n (i.e $e = 0$), and :**

A unit quantity is added to Q^{Opp} , under control "c", the transition probability is the following :

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, q^{bef}, q^{aft}, (q^{opp} + 1), e) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) \\ = \mathbb{P}(\{N_{n+1}^{Opp,+} - N_n^{Opp,+} = 1\} \cap \{N_{n+1}^{Opp,-} - N_n^{Opp,-} = 0\} \\ \cap \{N_{n+1}^{Same,+} - N_n^{Same,+} = 0\} \cap \{N_{n+1}^{Same,-} - N_n^{Same,-} = 0\}) \\ = \lambda_n^{Opp,+} \Delta_t (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t). \end{aligned}$$

Under control "s", nothing changes except that the order is cancelled and potentially inserted at the next step such as $Q_{n+1}^{Bef} = q^{bef} + q^{aft}$ and $Q_{n+1}^{Aft} = 0$.

A unit quantity is cancelled from to Q^{Opp} , under control "c", we differentiate between two cases:

1. **When $q^{opp} > 1$:**

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, q^{bef}, q^{aft}, (q^{opp} - 1), e) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) \\ = \mathbb{P}(\{N_{n+1}^{Opp,+} - N_n^{Opp,+} = 0\} \cap \{N_{n+1}^{Opp,-} - N_n^{Opp,-} = 1\} \\ \cap \{N_{n+1}^{Same,+} - N_n^{Same,+} = 0\} \cap \{N_{n+1}^{Same,-} - N_n^{Same,-} = 0\}) \\ = \lambda_n^{Opp,-} \Delta_t (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Same,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t). \end{aligned}$$

Under control "s", nothing changes except that $Q_{n+1}^{Bef} = q^{bef} + q^{aft}$ and $Q_{n+1}^{Aft} = 0$.

2. **When $q^{opp} \leq 1$, the price increase by one tick :**

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p + 1, q^{ins}, 0, q^{disc}, e) | U_n = (p, q^{bef}, q^{aft}, 1, e)) \\ = \lambda^{-Ask} \Delta_t (1 - \lambda^{+Ask} \Delta_t) (1 - \lambda^- \Delta_t) (1 - \lambda^+ \Delta_t) \mu_{n+1}(q^{disc}, q^{ins}). \end{aligned}$$

Under control "s", the last formula does'nt change.

A unit quantity is added to Q^{Same} , under control "c" we have :

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, q^{bef}, q^{aft} + 1, q^{opp}, e) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) \\ = \lambda_n^{Same,+} \Delta_t (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,-} \Delta_t). \end{aligned}$$

Under control "s", nothing changes except that $Q_{n+1}^{Bef} = q^{bef} + q^{aft} + 1$ and $Q_{n+1}^{Aft} = 0$.

A unit quantity is cancelled from Q^{Same} , under control "c", we distinguish again three cases :

1. **When $(q^{bef} > 1 \text{ and } q^{aft} \geq 0)$ or $(q^{bef} = 1 \text{ and } q^{aft} \geq 1)$:**

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, q^{bef} - 1, q^{aft}, q^{opp}, e) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) \\ = \lambda_n^{Same,-} \Delta_t (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t). \end{aligned}$$

We suppose that our order is executed when Q_n^{After} is consumed. Under control "s", nothing changes except that $Q_{n+1}^{Bef} = q^{bef} + q^{aft} - 1$ and $Q_{n+1}^{Aft} = 0$.

2. **When $q^{bef} = 0$ and $q^{aft} > 1$:**

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, 0, (q^{aft} - 1), q^{opp}, 1) | U_n = (p, q^{bef}, q^{aft}, q^{opp}, e)) \\ = \lambda_n^{Same,-} \Delta_t (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t). \end{aligned}$$

Under control "s", nothing changes except that $Q_{n+1}^{Bef} = q^{bef} + q^{aft} - 1$, $Q_{n+1}^{Aft} = 0$ and $Exec_{n+1} = 0$.

3. **When $q^{bef} + q^{aft} = 1$:**

$$\begin{aligned} \mathbb{P}^{(c)}(U_{n+1} = (p, q^{disc}, 0, q^{ins}, 1) | U_n = (p, 0, q^{aft}, q^{opp}, e)) \\ = \lambda_n^{Same,-} \Delta_t (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t). \end{aligned}$$

Under control "s", nothing changes except that $Exec_{n+1} = 0$.

Nothing happens in the limit order book with probability

$$(1 - \lambda_n^{Same,-} \Delta_t) (1 - \lambda_n^{Opp,+} \Delta_t) (1 - \lambda_n^{Opp,-} \Delta_t) (1 - \lambda_n^{Same,+} \Delta_t).$$

- For all the remaining cases we assume the transition probability neglectibe. We hence set it zero.

Remark. By taking in consideration the different cases and neglecting the terms with order strictly superior than 1 in Δ_t , we have for any control $i \in \{c, s\}$:

$$\begin{aligned} \sum_{\substack{\text{states } U_n \\ \text{states } U_{n+1}}} \int_{(\mathbb{N}_+)^2} \mathbb{P}^{(i)}(U_{n+1} | U_n) \mu_{n+1}(dq^{disc}, dq^{ins}) \\ \approx 1 + \lambda^{+Ask} \Delta_t + \lambda^{-Ask} \Delta_t + \lambda^+ \Delta_t + \lambda^- \Delta_t. \end{aligned}$$

Consequently, if $\lambda^{+Ask} \Delta_t + \lambda^{-Ask} \Delta_t + \lambda^+ \Delta_t + \lambda^- \Delta_t = o(1)$ (which is true when Δ_t is small), we end up with for any control $i \in \{c, s\}$:

$$\sum_{\substack{\text{states } U_n \\ \text{states } U_{n+1}}} \int_{(\mathbb{N}_+)^2} \mathbb{P}^{(i)}(U_{n+1} | U_n) \mu_{n+1}(dq^{disc}, dq^{ins}) \approx 1 \quad (16)$$

B Composition of market participants groups

High Frequency Traders

Name	NASADQ-OMX member code(s)	Market Maker	Prop. Trader
All Options International B.V.	AOI	Yes Yes	Yes Yes
Hardcastle Trading AG	HCT		
IMC Trading B.V	IMC, IMA		
KCG Europe Limited	KEM, GEL		
MMX Trading B.V	MMX		
Nyenburgh Holding B.V.	NYE		
Optiver VOF	OPV		
Spire Europe Limited	SRE, SREA, SREB		
SSW-Trading GmbH	IAT		
WEBB Traders B.V	WEB		
Wolverine Trading UK Ltd	WLV		

Table 3: Composition of the group of HFT used for empirical examples, and the composition of our “high frequency market maker” and “high frequency proprietary traders” subgroups.

Global Investment Banks

Name	NASADQ-OMX member code(s)
Barclays Capital Securities Limited Plc	BRC
Citigroup Global Markets Limited	SAB
Commerzbank AG	CBK
Deutsche Bank AG	DBL
HSBC Bank plc	HBC
Merrill Lynch International	MLI
Nomura International plc	NIP

Table 4: Composition of the group of Global Investment Banks used for empirical examples.

Institutional Brokers

Name	NASADQ-OMX member code(s)
ABG Sundal Collier ASA	ABC
Citadel Securities (Europe) Limited	CDG
Erik Penser Bankaktiebolag	EPB
Jefferies International Limited	JEF
Neonet Securities AB	NEO
Remium Nordic AB	REM
Timber Hill Europe AG	TMB

Table 5: Composition of the group of Institutional Brokers used for empirical examples.

C Extreme Imbalances

The decreasing slope at the right of the curve in Figure 7.a and 7.b when imbalance is highly positive (i.e $Q^{Same} \gg Q^{Opp} \approx 1$). In this situation, the order will be executed in general before the final time T_f without being followed by a price move (1) or will be executed at T_f and followed by a price move (2). In both cases, the final constraint ΔP is positive (see graph 12). Given that ΔP in case (2) is lower than ΔP in case (1) and situation (2) occurs more frequently when imbalance is highly positive, it is expected to find a decreasing slope at the right of the curve.

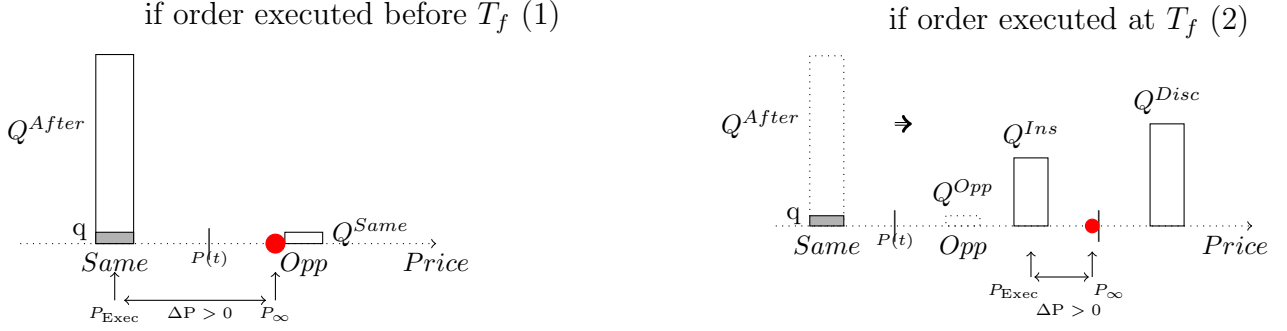


Figure 12: ΔP when Imbalance highly positive

At the opposite, when imbalance is highly negative (i.e $1 \approx Q^{Same} \ll Q^{Opp}$), the order will be executed as well in general before the maturity T_f and followed by a price change (1) or will be executed at f without being followed by a price move (2). In both cases, ΔP is negative (see graph 13). Given that ΔP in case (1) is greater than ΔP in case (2) and situation (1) occurs more frequently when imbalance is highly negative, it is expected to find an **increasing slope at the left of the curve**.

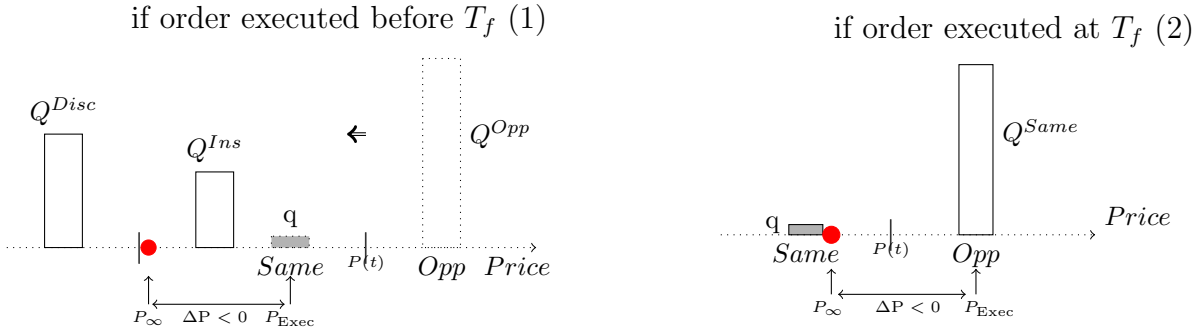


Figure 13: ΔP when Imbalance highly negative

The linear decreasing trend of the curve can also be explained by the expression of $\Delta P = \pm 0.5 + \frac{\alpha}{2} \times Imb$ specially when the imbalance effect isn't significant (far from the extreme points). We can also notice that the OC cases provide better result than the NC case when the imbalance is highly positive (i.e high adverse selection is present). This effect is reinforced when intensities depend on the imbalance. The linear decreasing of the curve is coherent with the empirical result in Figure 1.

D Influence of Non Constant Intensities

The framework (CONST) is a simplified version of the problem introduced first to shed light on some stylized facts like the adverse selection risk. The framework (IMB) is a more realistic and dynamic version of the problem because insertion and cancellation intensities depend on the current state of the markov chain U_n . Moreover, the inserted and discovered quantities Q^{Ins} and Q^{Disc} depend as well on the current state of the markov chain U_n . In this part, we comment the difference between theses two models.

The Figure 14.a represents the variation of $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ when the initial imbalance of the limit orderbook moves, under (CONST) and (IMB). In Figure 14.a, Blue points refer to the cases where it is better to stay in the orderbook at the beginning while red points refer to the cases where it is better to cancel the order at the beginning. We keep the same initial parameters of the Figure 7.

Let μ a control, we define $p_{U_0, \mu}^L$ the probability that the price P_n moves to the left starting from the initial state U_0 under the control μ . The quantity $p_{U_0, \mu}^L$ can be written :

$$p_{U_0, \mu}^L = \sum_{i=0}^f \sum_{\substack{\text{states } U_i \\ \text{under } \mu}} \mathbb{E}(1_{\{P_i - P_{i+1}=1\}})$$

Similarly, we define $p_{U_0, \mu}^R$ the probability that the price P_n moves to the right starting from the initial state U_0 under the control μ such as :

$$p_{U_0, \mu}^R = \sum_{i=0}^f \sum_{\substack{\text{states } U_i \\ \text{under } \mu}} \mathbb{E}_{U_0, \mu}(1_{\{P_{i+1} - P_i=1\}})$$

Thanks to the above equations, the quantities $p_{U_0, \mu}^R$ and $p_{U_0, \mu}^L$ can be directly computed by a forward algorithm that visits all the possible states of the markov chain U_n .

In Figure 14.b, we represent the variation of $p_{U_0, \mu}^L$ under (CONST) (i.e blue points) and (IMB) (i.e green points) when the initial imbalance moves. The circle radius is proportional to the number of times the price moves downward (i.e. in the adverse selection direction) : when the price moves downward frequently, the point size is huge. We represent as well the variation of $p_{U_0, \mu}^R$ under (CONST) (red points) and (IMB) (yellow points) when the initial imbalance moves. This time, the circle radius is proportional to the number of times the price moves upward (i.e. worst price). We keep the same initial parameters of the Figure 7.

Figure 14 shows **an imbalance effect**. When intensities depend on imbalance (IMB) it changes the transition probabilities by giving more weights to some events and less to others. For instance, assume imbalance is highly positive (cf. Figure 12). In the first case (cf. order executed before f) $\Delta P \approx 1$ (cf. Figure 12) while in the second case (cf. order executed at f) $\Delta P \leq \frac{1}{2}$ (it depends on how Q^{Disc} and Q^{Ins} are computed but it remains a sensible value of ΔP). As Figure 14.b shows that more weight is given to the second case when intensities depend on the imbalance (IMB). Consequently, it is expected to find $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ lower when intensities are variable (IMB). The same reasoning explains the shape of $\mathbb{E}_{U_0, \mu^*}(\Delta P | \text{Exec})$ when imbalance is highly negative.

E The influence of β

β is a predictability parameter that measures the sensitivity of event flows to the imbalance. When β is small, the orderbook is highly influenced by the imbalance and it moves quickly

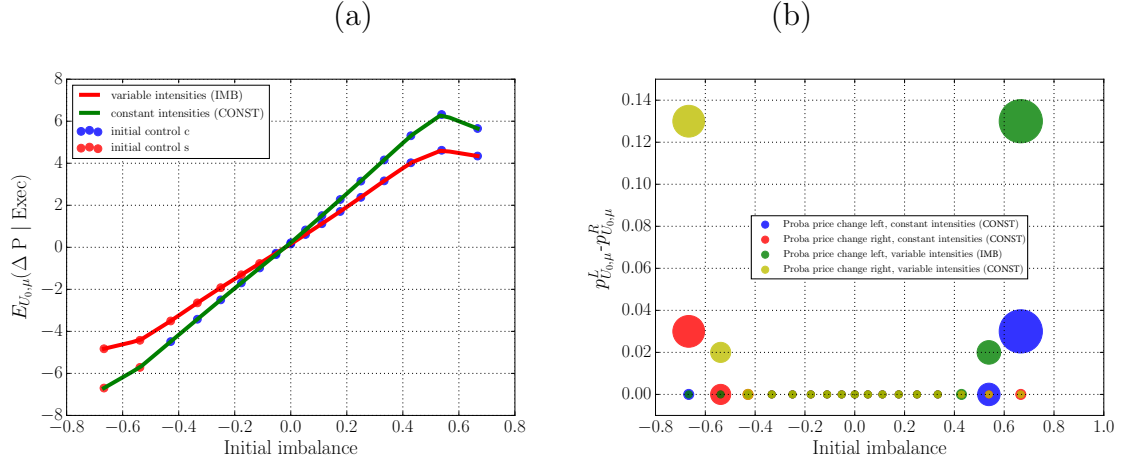


Figure 14: (a) $\mathbb{E}_{U_0, \mu^*}(\Delta P | Exec)$ move relative to initial imbalance under (CONST) and (IMB). (b) $p_{U_0, \mu}^L$ and $p_{U_0, \mu}^R$ move relative to initial imbalance under (CONST) and (IMB).

from a state to another. When β is great, orderbook moves are slower and less influenced by the imbalance.

The Figure 15.a represents the variation of $\mathbb{E}_{U_0, \mu^*}(\Delta P | Exec)$ when the initial imbalance of the orderbook moves, for different values of the parameter β . We keep the same initial parameters of the Figure 4.

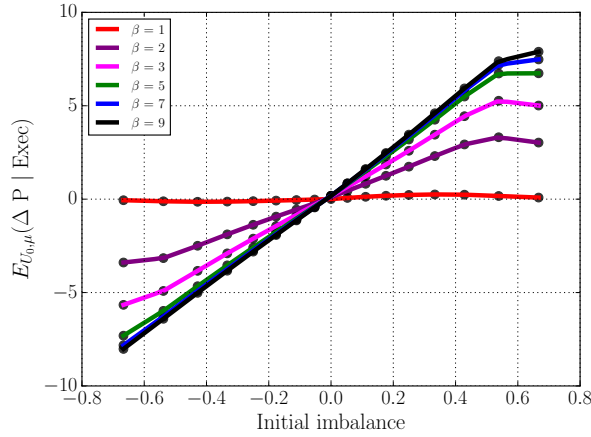


Figure 15: $\mathbb{E}_{U_0, \mu^*}(\Delta P | Exec)$ move relative to initial imbalance under (IMB) for different values of the parameter β

Figure 15 shows that when β is small $\mathbb{E}_{U_0, \mu^*}(\Delta P | Exec)$ is almost null and doesn't depend on the initial state of the order book. In fact, the order book moves so quickly that the influence of the initial state is insignificant. At the opposite, $\mathbb{E}_{U_0, \mu^*}(\Delta P | Exec)$ highly depend on the initial imbalance when β is great. We can see that β is a parameter that change the **slope of the curve**.